MGDA Variants for Multi-Objective Optimization

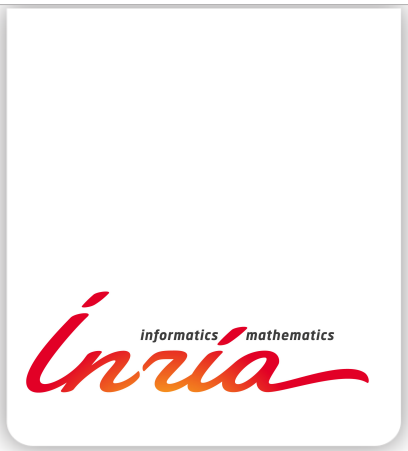
Jean-Antoine Désidéri

MGDA Variants for Multi-Objective Optimization

Jean-Antoine Désidéri

RESEARCH REPORT
N° 8068
September 17, 2012
Project-Team Opale

ISSN 0249-6399 ISRN INRIA/RR--8068--FR+ENG



MGDA Variants for Multi-Objective Optimization

Jean-Antoine Désidéri*

Project-Team Opale

Research Report n° 8068 — September 17, 2012 — 13 pages

Abstract: This report is a sequel to several publications in which a *Multiple-Gradient Descent Algorithm (MGDA)*, has been proposed and tested for the treatment of multi-objective differentiable optimization. Originally introduced in [2], the method has been tested and reformulated in [6]. Its efficacy to identify the Pareto front has been demonstrated in [7], in comparison with an evolutionary strategy. Recently, a variant, *MGDA-II*, has been proposed in which the descent direction is calculated by a direct procedure [4] based on a Gram-Schmidt orthogonalization process (*GSP*) with special normalization. This algorithm was tested in the context of a simulation by domain partitioning, as a technique to match the different interface components concurrently [3]. The experimentation revealed the importance of scaling, and a slightly modified normalization procedure was proposed ("*MGDA-IIb*"). In this new report, two novel variants are proposed. The first, *MGDA-III*, realizes two enhancements. Firstly, the *GSP* is conducted incompletely whenever a test reveals that the current estimate of the direction of search is adequate also w.r.t. the gradients not yet taken into account; this improvement simplifies the identification of the search direction when the gradients point roughly in the same direction, and makes the Fréchet derivative common to several objective-functions larger. Secondly, the order in which the different gradients are considered in the *GSP* is defined in a unique way devised to favor an incomplete *GSP*. In the second variant, *MGDA-IV*, the question of scaling is addressed when the Hessians are known. A variant is also proposed in which the Hessians are estimated by the Broyden-Fletcher-Goldfarb-Shanno (*BFGS*) formula.

Key-words: multi-objective optimization, descent direction, convex hull, Gram-Schmidt orthogonalization process, *BFGS* quasi-Newton method

* INRIA Research Director, Opale Project-Team Head

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Variantes de l'algorithme *MGDA* pour l'optimisation multiobjectif

Résumé : Ce rapport fait suite à plusieurs publications dans lesquelles on a proposé et testé un *Algorithme de Descente à Gradients Multiples (MGDA)* pour traiter les problèmes d'optimisation différentiable multiobjectifs. La méthode a été introduite originellement dans [2], et à nouveau formalisée dans [6]. Sa capacité à identifier le front de Pareto a été mise en évidence dans [7], en comparaison à une stratégie évolutionnaire. Récemment, une variante, *MGDA II*, a été proposée dans laquelle la direction de descente est calculée par une procédure directe [4] s'appuyant sur le processus d'orthogonalisation de Gram-Schmidt (*GSP*), effectué avec une certaine normalisation. On a testé l'efficacité de l'algorithme dans le contexte d'une simulation par partitionnement de domaine, comme technique pour raccorder concouramment les différentes composantes d'interface [3]. On a observé l'importance des facteurs d'échelle, ce qui a conduit a une légère modification de la procédure de normalisation ("*MGDA-IIb*"). Dans ce nouveau rapport, deux nouvelles variantes sont proposées. La première, *MGDA-III*, réalise deux améliorations. Premièrement, le *GSP* est exécuté incomplètement dès lors qu'un test révèle que l'estimation courante de la direction de recherche convient aussi aux vecteurs gradients qui n'ont pas encore été pris en considération; cette amélioration simplifie l'identification de la direction de recherche lorsque les gradients ont une direction commune approchée, et présente également l'avantage d'augmenter la dérivée de Fréchet commune aux fonctions objectifs. Deuxièmement, l'ordre dans lequel les differents gradients sont pris en compte dans le *GSP* est défini d'une manière unique conçue pour favoriser l'interruption rapide du *GSP*. Dans la seconde variante, *MGDA-IV*, la question du choix des échelles appropriées pour normaliser les gradients est abordée en supposant les hessiens connus. Une variante est également proposée dans laquelle les hessiens sont estimés par la formule de Broyden-Fletcher-Goldfarb-Shanno (*BFGS*).

Mots-clés : optimisation multiobjectif, direction de descente, enveloppe convexe, processus d'orthogonalisation de Gram-Schmidt, méthode *BFGS* de quasi-Newton

1 Introduction

We consider the simultaneous minimization or reduction of n objective-functions, $\{J_i(Y)\}$ $(i = 1, \ldots, n)$, assumed to be smooth (say $\mathcal{C}^2$) functions of the design-vector $Y \in \mathbb{R}^N$, where $n \leq N$.

Our analysis is developed to identify an appropriate direction of search ω to update the design vector from a given design-point Y^0, center of an open ball $\mathcal{B}$ in which the objective-functions are well-defined and smooth:

$$Y^1 = Y^0 - \rho\omega \qquad (\rho > 0, \text{ step-size}) \tag{1}$$

The above iteration is a descent step, if two conditions are met. Firstly, the vector $\omega \in \mathbb{R}^N$ should be such that

$$\forall i = 1, \ldots, n \, : \, \big(\nabla J_i(Y^0), \omega\big) > 0 \, . \tag{2}$$

Then, $-\omega$ is a descent direction common to all objective-functions. Secondly, the step-size ρ should be adjusted appropriately. In this report, we only focus on the first condition.

In [2] and [6], we have introduced the notion of "Pareto-stationarity":

Definition 1 (Pareto-stationarity)
The design-point Y^0 is said to be Pareto-stationary if there exists a convex combination of the gradients, $\nabla J_i(Y^0)$, that is equal to 0:

$$\exists \alpha = \{\alpha_i\} \; (i = 1, \ldots, n) \text{ such that : } \alpha_i \geq 0 \; (\forall i); \; \sum_{i=1}^{n} \alpha_i = 1; \; \sum_{i=1}^{n} \alpha_i \nabla J_i(Y^0) = 0 \, . \tag{3}$$

We have shown that Pareto-stationarity is a necessary condition to Pareto-optimality. Thereafter, we restrict ourselves to the case where the initial design-point Y^0 is not Pareto-optimal, thus not Pareto-stationary.

Clearly, the above condition (2), as it only involves scalar products, can be applied to projected gradients, in case of constrained minimization. More specifically, suppose that the active scalar constraints at $Y = Y^0$ are the following:

$$g_1(Y^0) = g_2(Y^0) = \cdots = g_K(Y^0) = 0 \, . \tag{4}$$

Then define the vectors

$$v_k = \nabla g_k(Y^0) \qquad (k = 1, \ldots, K) \tag{5}$$

normal to the constraint surfaces. Apply the *GSP* to these vectors to get the family $\{w_k\}$ $(k = 1, \ldots, K)$ of orthonormal vectors that collectively span the same subspace, and define the following projection matrix:

$$P = I_N - \sum_{k=1}^{K} [w_k] \, [w_k]^T \tag{6}$$

where the bracketed vector $[w_k]$ stands for the column-vector of its components viewed as a $N \times 1$ matrix, and the superscript ^{T} indicates transposition. Then, the forthcoming *MGDA* construction should be applied after replacing the original gradients, $\nabla J_i(Y^0)$, by their projections onto the subspace tangent to the constraint surfaces, that is, $P\nabla J_i(Y^0)$. This remark being made, without loss of generality, we are considering thereafter the unconstrained formulation.

In the original formulation of *MGDA* [2]-[6], the vector ω is defined as the minimum-norm element in the convex hull of the gradients:

$$\overline{\mathsf{U}} = \left\{ u \in \mathbb{R}^N \; / \; u = \sum_{i=1}^{n} \alpha_i \nabla J_i(Y^0) \, ; \; \alpha_i \geq 0 \; (\forall i); \; \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{7}$$

that is:

$$\omega = \mathsf{Argmin}_{u\in\overline{\mathsf{U}}} \|u\| \tag{8}$$

This definition is the most general and in particular, it is applicable whether the gradient vectors are linearly independent or not. The element ω can be identified by numerical minimization in the convex hull, which can be parameterized isomorphically to the hypercube $[0,1]^{n-1}$ (see [6]). This minimization can however be numerically delicate, and in fact, not necessary.

The convex hull can also be viewed as an affine structure, since:

$$\forall u \in \overline{\mathsf{U}} \,:\, u - u_n = \sum_{i=1}^{n} \alpha_i u_i - \left(\sum_{i=1}^{n} \alpha_i\right) u_n = \sum_{i=1}^{n-1} \alpha_i u_{n,i} \qquad (u_{n,i} = u_i - u_n) \tag{9}$$

Hence, $\overline{\mathsf{U}} \subseteq \mathcal{A}_{n-1}$ (or using affine-space notations, $\dot{\overline{\mathsf{U}}} \subseteq \dot{\mathcal{A}}_{n-1}$), where $\mathcal{A}_{n-1}$ is a set of vectors pointing to an affine sub-space $\dot{\mathcal{A}}_{n-1}$ of dimension at most $n-1$.

Let us examine these affine and vector structures, with the support of Fig. 1 drawn in the case $n = 3$. Here vectors are represented in the $\mathbb{R}^3$ affine space with a given origin O. The gradient vectors are here denoted $\{u_i\}$ $(i = 1, 2, 3)$. The convex hull of the gradients is the set of vectors of origin O pointing onto the triangle made of the 3 endpoints of $\{u_i\}$. This triangle lies in a plane (generally speaking a subspace of dimension at most $n-1$) denoted $\dot{\mathcal{A}}_2$. The orthogonal projection of O onto the plane $\dot{\mathcal{A}}_2$ is denoted $O^\perp$. The figure has been drawn in the case where $O^\perp \notin \dot{\overline{\mathsf{U}}}$.

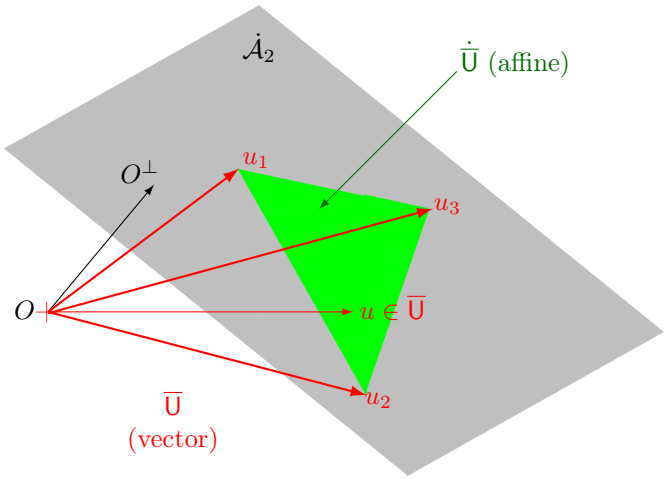


Figure 1: Affine and vector structures: here, three vectors $\{u_i\}$ $(i = 1, 2, 3)$ are considered to define the convex hull $\overline{\mathsf{U}}$, and u is an arbitrary element in $\overline{\mathsf{U}}$; all four vectors are associated with representatives of origin O; the endpoints of $\{u_i\}$ define the green triangle, $\dot{\overline{\mathsf{U}}}$, affine structure associated with the convex hull $\overline{\mathsf{U}}$; $\dot{\overline{\mathsf{U}}}$ is contained in the plane $\dot{\mathcal{A}}_2$; $O^\perp$ is the orthogonal projection of O onto $\dot{\mathcal{A}}_2$; the figure illustrates the case where $O^\perp \notin \dot{\mathcal{A}}_2$.

Now, in the particular case where $O^\perp \in \dot{\overline{\mathsf{U}}}$, or equivalently, $\overrightarrow{OO^\perp} \in \overline{\mathsf{U}}$, and since $\overrightarrow{OO^\perp} \perp \dot{\overline{\mathsf{U}}}$, $\omega = \overrightarrow{OO^\perp}$, and by orthogonality:

$$\big(u_i, \omega\big) = \|\omega\|^2 \qquad (\forall i) \tag{10}$$

This is a highly favorable situation in which all the Fréchet derivatives are equal.

The element ω being defined, the *MGDA* iteration generalizes the classical steepest-descent method [1] to multi-objective optimization by using the vector $-\omega$ as the direction of search. We have established that, certain weak provisions on the problem formulation being made, if the step-size is adjusted optimally, the iteration accumulates at a Pareto-stationary design-point. Whenever $\omega = 0$ (or, in a numerical procedure, when $\|\omega\| < TOL$), the current design-point is Pareto-stationary, and the optimization is interrupted. Hence, in [7], the efficacy of *MGDA* to identify the Pareto front has been demonstrated, and comparisons with an evolutionary strategy (*PAES*) have been made.

Recently, a variant, *MGDA-II*, has been proposed in which the descent direction is calculated by a direct procedure [4] based on a Gram-Schmidt orthogonalization process (*GSP*) with special normalization. This version is valid only when the gradient vectors are linearly independent, and it proceeds as follows. Due to the importance of scaling, user-supplied scaling factors $\{S_i\}$ $(i = 1, \ldots, n)$, are assumed given, and the following scaled gradients are defined:

$$J'_i = \frac{\nabla J_i(Y^0)}{S_i} \tag{11}$$

($S_i > 0$; e.g. $S_i = J_i$ for logarithmic gradients). The *GSP* is performed as follows:

- Set $u_1 = J'_1$
- For $i = 2, \ldots, n$, set:

$$u_i = \frac{J'_i - \sum_{k<i} c_{i,k} u_k}{A_i} \tag{12}$$

where:

$$c_{i,k} = \frac{\left(J'_i, u_k\right)}{\left(u_k, u_k\right)} \qquad (\forall k < i) \tag{13}$$

and:

$$A_i = \begin{cases} 1 - \displaystyle\sum_{k<i} c_{i,k} & \text{if nonzero} \\ \varepsilon_i & \text{otherwise } (\varepsilon_i \text{ arbitrary, but small}). \end{cases} \tag{14}$$

As a result of this construction, a new element ω is defined, as the minimum-norm element in the convex hull of the orthogonal vectors $\{u_i\}$ $(i = 1, \ldots, n)$

$$\omega = \sum_{i=1}^{n} \alpha_i u_i \tag{15}$$

in which the coefficients $\{\alpha_i\}$ are strictly positive and less than 1:

$$\alpha_i = \frac{1}{\|u_i\|^2 \sum_{j=1}^{n} \frac{1}{\|u_j\|^2}} = \frac{1}{1 + \sum_{j \neq i} \frac{\|u_i\|^2}{\|u_j\|^2}} < 1\,. \tag{16}$$

Due to the orthogonality of the family $\{u_i\}$ $(i = 1, \ldots, n)$, $\omega = \overrightarrow{OO^{\perp}}$, and (10) holds. This is illustrated by Fig. 2. Consequently

$$\left(J'_i, \omega\right) = \left(A_i + \sum_{k<i} c_{i,k}\right) \|\omega\|^2 = \|\omega\|^2 \tag{17}$$

by definition of the normalization constant A_i. In conclusion, the Fréchet derivatives of all objective-functions are equal.

It should be emphasized that in general the newly-defined element ω is distinct from the former, except in the particular case of two objective-functions ($n = 2$), when the gradient vectors form an obtuse angle.

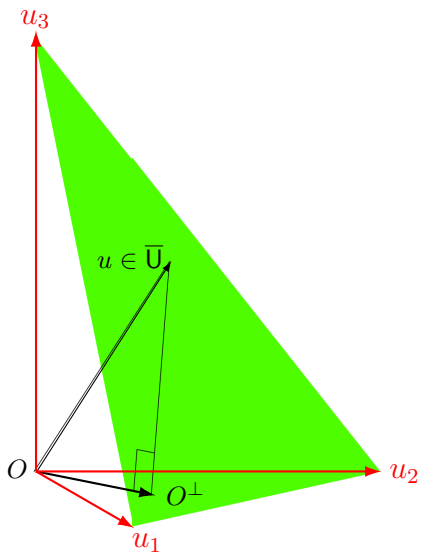


Figure 2: Construction of orthogonal vectors in *MGDA-II*

It was later observed that situations in which a normalization constant A_i for some i was negative was to be avoided. To do this, in *MGDA-II b*, the definition

$$A_i = 1 - \sum_{k<i} c_{i,k} \tag{18}$$

was maintained only if this number is strictly-positive. Otherwise, we have used the following modified scale:

$$S'_i = \left(\sum_{k<i} c_{i,k} \right) S_i \tag{19}$$

so that:

$$c'_{i,k} = \left(\sum_{k<i} c_{i,k} \right)^{-1} c_{i,k} \tag{20}$$

and

$$\sum_{k<i} c'_{i,k} = 1\,, \tag{21}$$

and $A_i = \varepsilon_i$, for some small ε_i. (This procedure was referred to as the "automatic rescale procedure".)

This has led us to the same formal conclusion: the Fréchet derivatives are equal; but the value is much larger, and (at least) one objective-function has been rescaled.

This variant was tested on a somewhat peculiar model problem of domain-partitioning, in which all objective-functions tend to 0, which results in a Pareto set restricted to a single point. In this application rather remote from the context for which *MGDA* has been devised, using logarithmic scaling of the gradients ($S_i = J_i$) and automatic rescale, *MGDA-II b* was found to converge satisfactorily; in fact, at a rate only twice less than the quasi-Newton method; additionally, the iteration indicated asymptotically an interesting trend to convergence acceleration [3].

Nevertheless, these developments have brought up some open questions to which this report brings certain answers. In particular, the following three:

1. Since the element ω provided by *MGDA-II* is in general different from the original one, how can we guarantee the convergence of *MGDA* to a Pareto-stationary design-point?

2. In which order should we arrange the gradients to perform the *GSP* ?
3. Can we devise an adequate scaling of the gradient inspired from quasi-Newton methods?

MGDA-III provides an answer to the first two questions, and *MGDA-IV* to the fourth.

2 Ordered and economical *GSP* : *MGDA-III*

The driving idea is that in case of numerous gradients, trends might emerge among them, permitting to account for the general direction of a subgroup by a unique vector in the orthogonal basis. Hence the *GSP* could be interrupted as soon as a direction is found to be a descent direction common to all objective functions while being constructed on the basis of only $I < n$ gradients. To achieve this purpose, in the following algorithm, at the stage of computing a new orthogonal vector, the gradient that is elected among those not yet accounted for, is the one for which the scalar product with the current estimate of the element ω is algebraically smallest. In this way, it is the vector for which the construction so far is the least satisfactory. Thus, computational economy is achieved through the specification of the ordering in which the gradients are used to perform the *GSP* , with the expectation of a rapid interruption of the process. Further comments on the expected gain in efficiency will be made subsequently.

2.1 Algorithm

Again, one starts from scaled gradients, $\{J'_i\}$, and duplicates, $\{g_i\}$,

$$J'_i = \frac{\nabla J_i(Y^0)}{S_i} \quad (S_i : \text{ user-supplied scale}) \tag{22}$$

and proceeds in 3 steps: A,B and C.

A: Initialization

- Set[1]
$$u_1 := g_1 = J'_k \ / \ k = \mathsf{Argmax}_i \min_j \frac{(J'_j, J'_i)}{(J'_i, J'_i)} \tag{23}$$
- Set $n \times n$ lower-triangular array $c = \{c_{i,j}\}$ $(i \geq j)$ to 0.[2]
- Set, conservatively, $I := n$.[3]
- Assign some appropriate value to a cut-off constant $a : (0 \leq a < 1)$.

B: Main *GSP* loop; for $i = 2, 3, \ldots,$ (at most) n, do:

1. Calculate the $i-1$st column of coefficients:
$$c_{j,i-1} = \frac{(g_j, u_{i-1})}{(u_{i-1}, u_{i-1})} \qquad (\forall j = i, \ldots, n) \tag{24}$$
and update the cumulative row-sums:
$$c_{j,j} := c_{j,j} + c_{j,j-1} = \sum_{k<i} c_{j,k} \qquad (\forall j = i, \ldots, n) \tag{25}$$

[1] The choice made for g_1 will be justified afterwards.
[2] The main diagonal of array c is to contain cumulative row-sums.
[3] The integer $I \leq n$ is the expected number of computed orthogonal basis vectors.

2. Test:

   – If the following condition is satisfied

$$c_{j,j} > a \qquad (\forall j = i, \ldots, n) \tag{26}$$

   set $I := i - 1$, and interrupt the *GSP* (go to 3.).

   – Otherwise, compute next orthogonal vector u_i as follows:

   # *Identify index* $\ell = \mathsf{Argmin}_j \ \{c_{j,j} \ / \ i \le j \le n\}$.[4]

   # *Permute information associated with* i *and* ℓ :

$$g\text{-vectors: } g_i \rightleftharpoons g_\ell$$

   rows i and ℓ of array c and corresponding cumulative row-sums

$$c_{i,i} \rightleftharpoons c_{\ell,\ell}$$

   # *Set* $A_i = 1 - c_{i,i} \ge 1 - a > 0$ *and calculate*[5]

$$u_i = \frac{g_i - \sum_{k<i} c_{i,k} u_k}{A_i} \tag{27}$$

   # *If* $u_i \neq 0$, *return to 1. with incremented* i; *otherwise:*

$$g_i = \sum_{k<i} c_{i,k} u_k = \sum_{k<i} c'_{i,k} g_k \tag{28}$$

   where the $\{c'_{i,k}\}$ *are calculated by backward substitution.*
   Then, if $c'_{i,k} \le 0 \ (\forall k < i)$*:*

   Pareto-stationarity detected: STOP MGDA iteration;

   otherwise (exceptional ambiguous case):

   STOP GSP *; compute* ω *according to original definition and go to C.*

3. Calculate ω as the minimum-norm element in the convex hull of $\{u_1, u_2, \ldots, u_I\}$:[6]

$$\omega = \sum_{i=1}^{I} \alpha_i u_i \neq 0 \tag{29}$$

   where:

$$\alpha_i = \frac{1}{\|u_i\|^2 \sum_{j=1}^{I} \frac{1}{\|u_j\|^2}} \frac{1}{1 + \sum_{j \neq i} \frac{\|u_i\|^2}{\|u_j\|^2}} \,. \tag{30}$$

C: If $\|\omega\| < TOL$, STOP *MGDA* iteration; otherwise, perform descent step and return to B.

2.2 Properties

Case I = n. In this case, the *GSP* is performed completely, and the algorithm is equivalent to the former *MGDA-II* with the enhancement that the rescale of the *b-version* is never necessary, since the specified ordering implies that

$$\forall i \ : \ A_i \ge 1 - a > 0 \,. \tag{31}$$

[4] Note that necessarily $c_{\ell,\ell} \le a < 1$.

[5] $c_{i,i}$ = former-$c_{\ell,\ell} \le a$; g_i = former-$c_{\ell,k}$.

[6] Note that ω is calculated on the basis of a smaller number of gradients if $I < n$; here all computed $u_i \neq 0$, and $0 < \alpha_i < 1$.

Case I < n (incomplete *GSP*). Here, the Fréchet derivatives satisfy different bounds according to two subcases:

- First I Fréchet derivatives:

$$\big(g_i, \omega\big) = \big(u_i, \omega\big) = \|\omega\|^2 > 0 \qquad (\forall i = 1, \ldots, I) \tag{32}$$

- Subsequent ones ($i > I$):

By construction, the vectors $\{u_1, u_2, \ldots, u_I\}$ are orthogonal, and ω is given by (29), so that:

$$g_i = \sum_{k=1}^{I} c_{i,k} u_k + v_i \tag{33}$$

where $v_i \perp \{u_1, u_2, \ldots, u_I\}$. Consequently,

$$\big(g_i, \omega\big) = \sum_{k=1}^{I} c_{i,k} \big(u_k, \omega\big) = \sum_{k=1}^{I} c_{i,k} \|\omega\|^2 = c_{i,i} \|\omega\|^2 > a \|\omega\|^2 > 0 \tag{34}$$

2.3 A posteriori justification of the choice of g_1

At initialization, we have set

$$u_1 := g_1 = J'_k \ / \ k = \mathsf{Argmax}_i \min_j \frac{\big(J'_j, J'_i\big)}{\big(J'_i, J'_i\big)} \tag{35}$$

This was equivalent to maximizing $c_{2,1} = c_{2,2}$, that is, maximizing the least cumulative row-sum, at first estimation. This makes at start, the worst case less severe. One anticipates that the favorable situation for which all cumulative row-sums are positive (or $> a$), is more likely.

2.4 Expected benefits

According to the section above, the specified ordering has been devised to permit the *GSP* to be performed incompletely. When gradients exhibit a general trend, ω is found in fewer steps and this realizes a computational economy.

Secondly, the rescale procedure has been circumvented.

Thirdly, an incomplete *GSP* results in an element ω of larger norm since it realizes the minimization in a smaller subset, namely the convex hull of an incomplete orthogonal basis. This corresponds to larger directional derivatives, since

$$\big(g_i, \frac{\omega}{\|\omega\|}\big) = \|\omega\| \ \text{ or } a \|\omega\| \tag{36}$$

and to the greater efficiency of the subsequent *MGDA* descent step.

3 Using Hessians to better scale the gradients: *MGDA-IV*

3.1 Adressing the question of scaling when Hessians are known

In single-objective optimization, when both gradient and Hessian are known, Newton's method is the most effective unless additional information is provided.

For the optimization of the objective $J_i(Y)$ alone, Newton's method writes:

$$Y^1 = Y^0 - p_i \tag{37}$$

where the vector p_i is given by the solution of the system:

$$H_i p_i = \nabla J_i(Y^0) \tag{38}$$

where H_i is the Hessian matrix of objective function J_i at $Y = Y^0$. Hence the preconditioning by the inverse Hessian realizes a form of optimal scaling. However, in general, the vector p_i is not parallel to the gradient itself. Thus to ensure that the iteration remains a descent step, only its projection should be retained.

Thus, we propose to split the vector p_i into orthogonal components

$$p_i = q_i + r_i \tag{39}$$

where:

$$q_i = \frac{\big(p_i, \nabla J_i(Y^0)\big)}{\|\nabla J_i(Y^0)\|^2} \nabla J_i(Y^0) \tag{40}$$

and $r_i \perp \nabla J_i(Y^0)$, and to define the scaled gradient as follows:

$$J_i' = q_i \tag{41}$$

We define *MGDA-IV* as *MGDA-III* applied to the gradients scaled as above. With the previous notations, this is equivalent to defining the scaling constant S_i as follows:

$$S_i = \frac{\|\nabla J_i(Y^0)\|^2}{\big(p_i, \nabla J_i(Y^0)\big)} \tag{42}$$

3.2 *BFGS*-inspired variant: *MGDA-IV b*

When the Hessians are not known exactly, they can be approximated by the *BFGS* iterative estimate:

$$(\forall i = 1, \ldots, n) \quad \widetilde{H}_i^{(0)} = Id \tag{43}$$

$$\widetilde{H}_i^{(k+1)} = \widetilde{H}_i^{(k)} - \frac{1}{{s^{(k)}}^T \widetilde{H}_i^{(k)} s^{(k)}} \widetilde{H}_i^{(k)} s^{(k)} {s^{(k)}}^T {\widetilde{H}_i^{(k)}}^T + \frac{1}{{z_i^{(k)}}^T s^{(k)}} z_i^{(k)} {z_i^{(k)}}^T \tag{44}$$

in which k is the *MGDA* iteration index, and:

$$s^{(k)} = Y^{(k+1)} - Y^{(k)} \tag{45}$$

$$z_i^{(k)} = \nabla J_i(Y^{(k+1)}) - \nabla J_i(Y^{(k)}) \tag{46}$$

4 Conclusion

The different proposed variants of the MGDA are summarized in Table 1 with the indication of the major enhancements realized by each version, and references to publications and test-cases considered.

At this stage, further research is necessary to experiment *MGDA-III* and *MGDA-IV* and assess in practical engineering applications the improvements actually achieved by the theoretical enhancements. In particular, test-cases involving more complex Pareto fronts corresponding to a larger number n of objective functions, and possibly discontinuities, should be examined. Scaling with exact or approximate Hessians is a very promising option, but it should be verified cautiously also. Additionally, the question of step-size adjustment should also be investigated thoroughly.

Algorithm	Novel elements	Properties	Tested cases
MGDA	- General constructive principle related to minimum-norm element in convex hull of gradients [2] [6]	- CV proof to Pareto stationary designs - Insensitive to Pareto front convexity	- Multiple quadratics - Fonseca testcase (non-convex Pareto front; comparison *MGDA* vs *PAES*) [7] - DDM for Poisson pb. [3] [5]
b-version	- Meta-model assisted gradient computation [8]	- CV requires a few database enrichments	- Eulerian flow about wing [8] - Navier-Stokes duct flow [8]
MGDA-II	- Direct computation of descent direction by *GSP* [4] [3]	- Requires linearly independent gradients - Modified definition of descent direction, and Pareto-stationarity test necessary - $n!$ possible orderings	- DDM for Poisson pb. (scaling essential; verified CV to unique Pareto-stationary solution) [3]
b-version	- Automatic gradient rescale when normalization coefficient is found <0	- More efficient (larger directional derivatives)	- (id.)
MGDA-III	- Specific ordering in *GSP* - Incomplete *GSP* - Resort to standard *MGDA* when Pareto-stationarity test ambiguous	- Not limited to linearly-independent gradients - Even larger directional derivatives - Pareto-stationary accumulation points	
MGDA-IV	- Scaling inspired from Newton's method using Hessians	- Step-size of order unity expected to be appropriate	
b-version	- Uses *BFGS* approximations to Hessians		

Table 1: Variants of *MGDA* with details on progressive enhancements

References

[1] *Gradient descent*, WikipediA: The Free Encyclopedia , `http://en.wikipedia.org/wiki/Gradient_descent`.

[2] Jean-Antoine Désidéri, *Multiple-Gradient Descent Algorithm (MGDA)*, Research Report 6953, INRIA, 2009, `http://hal.inria.fr/inria-00389811`.

[3] ———, *Application of MGDA to domain partitioning*, Research Report 7968, INRIA, May 2012, `http://hal.inria.fr/hal-00694039`.

[4] ———, *MGDA II: A direct method for calculating a descent direction common to several criteria*, Research Report 7422, INRIA, April 2012, `http://hal.inria.fr/hal-00685762`.

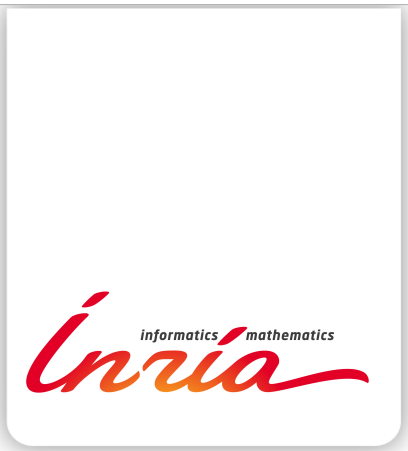
[5] ———, *Multiple-Gradient Descent Algorithm for Multiobjective Optimization*, Proc. European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS), September 10-14, 2012, Vienna, Austria, 2012, To appear in Skopus.

[6] ———, *Multiple-gradient descent algorithm (MGDA) for multiobjective optimization*, Comptes rendus - Mathématique **350** (2012), no. 5-6, 313–318, `http://dx.doi.org/10.1016/j.crma.2012.03.014`.

[7] Adrien Zerbinati, Jean-Antoine Désidéri, and Régis Duvigneau, *Comparison between MGDA and PAES for Multi-Objective Optimization*, Research Report 7667, INRIA, June 2011, `http://hal.inria.fr/inria-00605423`.

[8] ———, *Application pf the Multiple-Gradient Descent Algorithm (MGDA) and Metamodels to a Multiobjective Problem in Aerodynamics*, Proc. European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS), September 10-14, 2012, Vienna, Austria, 2012, To appear in Skopus.

Contents



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399